International Journal for Modern Trends in Science and Technology
Volume 9, Issue 08, pages 89-94
ISSN: 2455-3778 online
Available online at: http://www.ijmtst.com/vol9issue09.html
DOI: https://doi.org/10.46501/IJMTST0909071

Machine Learning Approaches for Prediction and Prevention of Cyber Attacks for Cyber Security

Pranali R Landge[1] | Dr. Swati S.Sherekar[2]

[1]Department of Computer Science & Engineering, Sant Gadge Baba Amravati University, Amravati,
[2]Department of Computer Science & Engineering, Sant Gadge Baba Amravati University, Amravati, Maharashtra, India

To Cite this Article
Pranali R Landge and Dr. Swati S.Sherekar. Machine Learning Approaches for Prediction and Prevention of Cyber Attacks for Cyber Security, International Journal for Modern Trends in Science and Technology, 2023, 9(09), pages. 89-94.https://doi.org/10.46501/IJMTST0909071

Article Info
Received: 30 August 2023; Accepted: 24 September 2023; Published: 27 September 2023.

ABSTRACT

The cost of data violation will increase due to the current fast digitization. Cyber risks caused by hackers and other online criminals frequently result in a lack of data security, which in turn causes large financial losses and a bad reputation for the company. Over the past several years, there has been a steady increase in the number of cyberattacks committed against growing enterprises. Since human analysis of cyber threat discovery and assistance requires a lot of time, money, and is expensive and prone to error, it is an impractical way to detecting cyber threats and attacks. Intelligent and automated help based on machine learning algorithms is needed to successfully prevent, identify, and respond to cyber-attacks. As a result, this calls for sophisticated automated help.

KEYWORDS:: Cyber threat, Machine Learning

1. INTRODUCTION

The surge in frequency and severity of cyber-attacks over the past few years is a significant challenge for many of the world's fastest-growing businesses. It is not a realitic method to manually search for cyber threats and attacks since doing so demands a significant amount of time and money, is an error-prone procedure, and is expensive. It is vital to make use of sophisticated automated assistance processes that make use of machine learning techniques such as Support Vector Machine and linear discriminant analysis in order to successfully avoid, recognize, and respond to cyber threats. These kinds of procedures are very necessary.

The financial repercussions of a cyberattack are becoming increasingly severe: Malware and other forms of web-based assaults are two of the most expensive types of cyberattacks. The percentage of ransomware assaults has increased from 13% to 27%. When it comes to information technology security investments, 30% are allocated to the network layer, while only 20% are allocated to the application layer [1]. In 2017, the United States had an average cost of cybercrime of $21 million, which is an increase over the average cost of $17 million in 2016. The existing system for cyber security takes about the same amount of time that other systems do to recognize threats, prevent them, and respond to them.

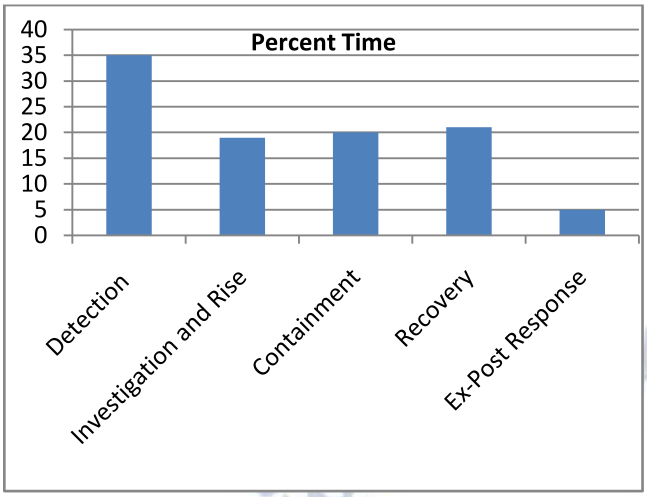

Figure 1: Average time to detect, prevent and respond to cyber threat

Declare that the broad term "cyber security" applies to any techniques and tools used to keep an eye on, block, or modify unauthorized access to, abuse of, or denial of service on computer networks. This also applies to the propensity for network-accessible essential infrastructure and restricted resources to be made available to them [2].

Security risks must be identified and addressed in two different ways in order to fight against cyberattacks: proactively, before an attack occurs, and reactively, after an attack has already occurred. The majority of the time, statistical techniques, association rules, machine learning algorithms, and evolutionary algorithms may be used to identify and predict attacks. To spot and stop risky behavior, traffic analysis is performed, which operates quite similarly to the majority of attack prevention systems [3]. This may be seen in image 2 below:

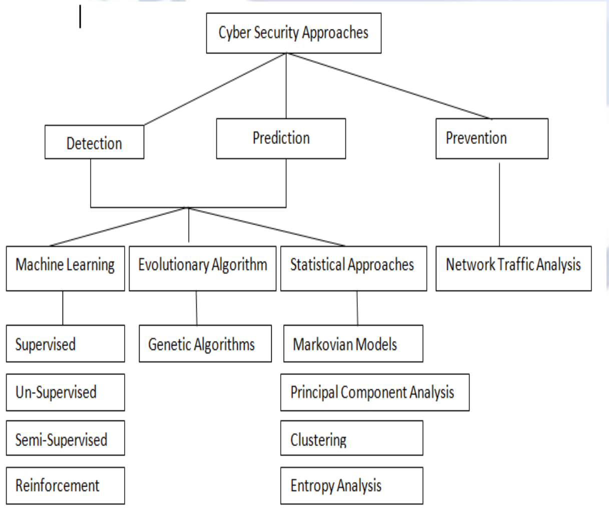

Figure 2: Cyber Security Approaches

2. RELATED WORK:

The field of intelligent cyber security systems has given rise to the development of a variety of strategies and protocols for the purpose of defending against cyberattacks. The work that is now available for intelligent cyber security-based detection systems is outlined in Table 1.

Author	Characteristics	Methodology /Techniques used	Advantages	Limitations
Carla Maria Sayan(2017) [1] IEEE 2nd Int'l Workshop	focuses on the architecture and design process for a smart assistant that can give advice on how to resolve cyber security problems and intelligently support a human security specialist.	Create Domain Specific Data model, Identify Features. Identify Implement and Trained the system. Implemented artificial intelligent using machine learning methods.	ICSA flawlessly and protectively detect attacks and vulnerabilities and stop attacks.	Recommending based on specific domain knowledge
Guang Xiang et al.(2011) [4] ACM Transaction	Eight novel features have been proposed as part of a comprehensive feature-based approach that uses ML techniques to exploit the DOM of HTML documents, search engines, and third-party services to identify phishing attacks. This approach is known as CANTINA+. Two filters were carefully chosen by the	Utilizing time-based and randomized evaluation techniques. Different tools and techniques are employed in this case, Bayesian Networks, Logistics Regression, Support Vector Machines (SVM) ,J48 Decision Trees, Random Forests (RF), and Adaboost are including in this case.	Constant lyidentify new phishing attacks and under the acceptable level controls the FPR (False Positive Rate)	Fail to detect legitimate domain and host fishing attacks on servers.

	authors to help minimize FP and speed up runtime. A good anti-phishing solution is CANTINA+.			
Naseer R. Sabar et al. (2018) [5] IEEE Access	The most difficult problem in big data cybersecurity is malware detection, because the majorities of malware detection techniques are designed primarily for small datasets and are therefore unable to handle huge data in a timely manner. This paper provides a new bi-objective hyper-heuristic framework for SVM configuration optimization to overcome the aforementioned problems.	The hyper heuristic framework & the SVM make up the two components of the methodology. Malware detection approaches and tools include detection for Signature based, Pattern based support vector machines, and cloud-based detection.	The bi-objective optimization problem can be resolve successfully and potentially.	According to the study, the proposed framework is extremely successful, although it is not necessarily superior when compared to equals and other algorithms.
R.Vinay kumar et al.(2019) [6] IEEE Access	IDS focuses on automatically and appropriately identifying and classifying cyberattacks at the host and network levels. Represent Text	Python was used to implement each experiment on an Ubuntu 14.0.4 LTS system. Utilizing Scikit-learn, all conventional machine learning	able to more correctly detect threats by outperforming HIDS and NIDS's currently employed	Complex Deep Neural Netwok structures have a high computational cost. Therefore they were not trained
	Representation Techniques and Scalable Computing Architecture	techniques were implemented. TensorFlow4, a GPU-enabled backend, was used to create deep neural networks (DNNs), together with the higher-level Keras framework.	traditional ML classifiers	in this study utilizing benchmark Intrusion Detection System (IDS) datasets.
AthorSubroto et al. (2019) [7] Springer Open Access	Discuss the recently introduced algorithmic model that predicts cyber hazards using statistical ML and big data analytics from social media. Managers of both public and commercial organisations might develop useful strategies for reducing cyber risks to key infrastructures using the study's findings.	Using R software and a local MySQL database, descriptive analysis, pyramid analysis, and prediction analysis using algorithms were performed.	With an accuracy rate of 96.73, ANN outperforms the competition. Create efficient plans for lowering the cyber risks to vital infrastructure.	Delays in the amount of time between the discovery of a vulnerability and its utilization in cyber risk management.

Table 1: Literature review

3. CYBER SECURITY APPROACHES

The Intrusion Detection System, sometimes referred to as IDS, is primarily used to identify a range of potentially hostile computer and network behavior. An "intrusion" is defined as an unlawful act that harms a computer system. It is feasible to shield computer systems against intrusions that might jeopardize their dependability, accessibility, or privacy by using an intrusion detection system (IDS). Instead of using signature-based databases, the Advanced Intelligent Detection System (AIDS) generates a threat signal anytime it notices unexpected behavior. Because it observes user activity and searches for anything unusual, AIDS can spot zero-day assaults. When testing for AIDS, there is a high likelihood of getting a false positive result. These traits allow AIDS to be divided into the following five subclasses: 1. based on the information gathered 2. based on recurrent themes 3. controlled by laws 4.State-based 5.Heuristic-basedThe following are the three main AIDS treatments:

1. Statistics-based AIDS
2. Knowledge-based AIDS
3. Machine learning-based AIDS [8]

A. DETECTION BY MACHINE LEARNING APPROACH

Iqbal H. Sarker et al. (2020) [9] emphasize cyber security data science (CDS), which entails security data attention, applying ML algorithms to assess cyber dangers, and ultimately attempting to optimize cyber security operations. The strategies used include hybrid approaches, stateful protocol analysis-based approaches, IDS based on anomalies, IDS based on signatures, and IDS based on signatures. Smart cyber security systems and services utilize supervised and unsupervised learning, SVM, and KNN techniques to intelligent decision-driven system building.

i) Supervised Learning Approach

Training and examination are the two primary stages that comprise supervised learning. Through the use of supervised learning, a classifier may be educated to detect the innate connection that exists between the value of the input data and the value of the labeled output. Labeled data are required for supervised learning. During testing, the trained model is applied to categorize unknown data as belonging to the intrusion class or the normal class, respectively. The constructed classifier then makes a prediction about the possible class that the input data may belong to by making use of a collection of attribute significances. A few examples of classification algorithms are neural networks, naive bayes, nearest neighbor, support vector machines (SVM), decision trees, and rule-based systems. Neural Networks is another classification strategy. Each method makes use of a different learning process in order to construct a classification model.

ii) Un-Supervised Learning Approach

One type of approach that may be applied in machine learning is an unsupervised learning strategy. Without using any sort of class labels, the unsupervised learning approach analyzes input datasets and extracts usable data from them. The data points input into the system are seen as a collection of random variables throughout the unsupervised learning process. Unsupervised learning refers to a type of learning when no labels are used and no labeled data is present. Instead, the data is automatically divided into several groups by the learning process. The technique of unsupervised learning, which entails training a model with unlabeled data, enables the identification of intrusions.

iii) Semi-Supervised Learning Approach

Learning that is semi-supervised may be broken down into two distinct categories: supervised learning and unsupervised learning. Supervised learning makes use of labeled data, whereas unsupervised learning does not. When semi-supervised learning and a limited number of labeled data are combined, the amount of time and money required for the IDSs to function is significantly reduced. This is promising because numerous factors that might affect an Intrusion Detection System could result in tagged data that is irregular or infrequent. There have been many further semi-supervised learning systems proposed, some of which include the semi-supervised Support Vector Machine, EM-based algorithms, graph-based methods, self-training, co-training, and boosting-based techniques [10].

iv) Reinforcement Learning Approach

In the field of pattern recognition, the strategy of reinforcement learning is an absolutely necessary component. The reinforcement learning technique enables software agents to learn from their interactions with their surroundings and select the most advantageous action to take in order to maximize their potential for long-term rewards [11]. The reinforcement learning methodology allows for a sensor node, also known as a software agent, to be educated by interacting with its surrounding environment. Analyzing problems with learning control and finding solutions to difficulties in sequential reinforcement learning are two applications of the Markov Decision Process (MDP) model. These are difficult issues to solve when using supervised learning. The difficulty depicted in figure 3 below pertains to reinforcement learning [12].

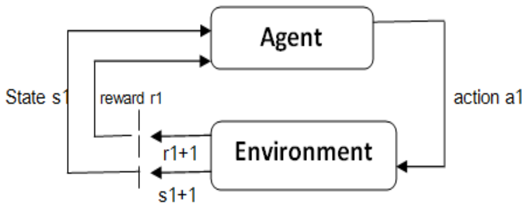

Figure 3: Reinforcement learning.

B. CYBER ATTACK PREDICTING APPROACH

The significant goal of developing a defensive system and their security capabilities [13] is to fulfill the key objective of cyber-attack prediction. E-correlator is an entropy-based alert correlation technique that was developed by M. GhasemiGol and colleagues [14] to make it simpler to evaluate a large collection of alerts while guaranteeing that no information is lost between the allied and the original raw alerts. This was accomplished by ensuring that there is no information loss between the allied and the raw alerts themselves. The E-correlator system receives raw alerts as input, which then results in the generation of a hyper-alerts graph. The approach models [15] the process of predicting attacks by making use of significant events that occur outside of an event window and are then put to use in the construction of an attack tree. This multistage attack detection and prediction provided findings with an accuracy of 95% in both cases. Attack trees can be constructed to offer a comprehensible output while also simulating the decision-making process of an attacker. The process of recreating assault methods can benefit from using attack trees.

C. CYBER ATTACK PREVENTING APPROACH

A proactive strategy that can quickly identify and address possible threats is necessary for the prevention of cyberattacks. The majority of methods are reactive, which means that they can only work if the hit zone has sustained sufficient damage. Many different forms of intrusion anticipatory systems have been expected as a potential way to increase the safety of cyberspace. Growing businesses require greater security and more advanced technologies to quickly safeguard their systems from attackers' intrusions. One of the needs that must be addressed in order to attain defensible security is having a product or approach that is not only affordable but also adaptable and scalable. This essay's main objective is to demonstrate the use of a real-time approach to identify and neutralize threats [16].

4. ANALYSIS& DISCUSSION

The most important strategies for recognizing, anticipating, and preventing cyberattacks have been discussed in this article. The discovery of patterns in security logs prior to an attack is one use of machine learning in the field of cyber security. Other applications include the detection of virus attacks, anomaly detection, cluster-based user profiling, and other similar uses. In the field of data analytics, one well-known method is known as machine learning. The vast bulk of research on vulnerabilities and defensive strategies may be placed into either of these two buckets. i.e., an exhaustive investigation into new methods and technology for providing temporary security and protection. There are three primary groups of cyber analytics that fall under the umbrella of an intrusion detection system: 1) Based on a signature 2) A sort based on anomalies, and 3) a hybrid variety. Throughout this article, we have discussed the significance of machine learning, as well as its various subtypes and ML approaches. An application of the semantic Machine Learning (ML) approach for the cyber risk detection and protection monitoring system reveals both the operational model and the functionality of the system. Data confidentiality is maintained when appropriate semantic machine learning algorithms and machine learning processors are utilized for purposes of data identification and protection [17]. Utilizing machine learning algorithms to quickly guard against more current assaults is one potential method that may be utilized to put a halt to cyberattacks. Both the temporal complexity of machine learning algorithms and the real-time environment of the pace at which cyber-attacks are detected need to be taken into consideration. There is a possibility that future research may concentrate on reducing processing costs and increasing detection rates using distributed approaches [18].

5. CONCLUSION

An intelligent automated decision support system that makes use of automatic cyber security tools and techniques based on machine learning algorithms generated a beneficial outcome in its attempts to minimize the severity of cyber threats, avoid the occurrence of event behavior, and prescribe measures. Using training data, the approaches of machine learning are able to get a wide variety of pattern matching skills. The use of ML techniques involves the identification of cyber-attacks through the activation of danger signals whenever observed behavior deviates from what is considered to be normal behavior. Additionally, the ML technique shortened the amount of time needed for prediction and prevention in relation to network monitoring and threat detection. Using a variety of ML approaches, an attack prevention system may be developed rather quickly. An effective intrusion detection strategy (IDS) should be able to find many forms of attacks, specifically incorporating an invasion that integrates avoidance tactics and high protection security against activities.

Conflict of interest statement

Authors declare that they do not have any conflict of interest.

REFERENCES

[1] C. Sayan, G. Ball and S. Hariri, " Cyber Security Assistant: design overview", IEEE 2nd International Workshop on Foundations and Applications of Self Systems, Tucson, AZ, 2017.

[2] . Buczak and E. Guyen, " A survey of data mining and machine learning methods for cyber security intrusion detection", IEEE communications Surveys & Tutorials, Vol. 18, no. 2, 2016, PP.1153-1176.

[3] D. E. Denning, " Framework and Principles for active cyber defence", Computer & Security, vol. 40, 2014, PP. 108-113

[4] Guang Xiang, Jason Hong, Carolyn P. Rose, Lorrie Cranor, " CANTINA+:A Feature- Rich Machine Learning Framework for Detecting Phishing Websites", ACM Transaction on Information and System Security, September 2011, Vol. 14, No. 2, Article 21, PP. 1-28.

[5] Nasser R. Sabar, Xun Yi, Andy Song, "A Bi-objective Hyper-Heuristic Support Vector Machines for Big Data Cyber-Security", IEEE Access, 2018, Volume 6, PP. 10421-10431

[6] R. Vinaykumar, MamounAlazab, K. P. Soman, PrabhaharanPoornachandran, Ameer AL-Nemrat, SitalakshmiVenkatraman," Deep Learning Approach for Intelligent Intrusion Detection System", IEEE Access, 2019, volume 7, PP. 41525-41550

[7] ArthorSubroto, AndriApriyana, " Cyber Risk Prediction Through Social media big data analytics and statistical machine learning", Springer Open Access Journal on Big Data, 2019, PP 1-19

[8] AnsamKhraisat, Iqbal Gondal, Peter Vamplew, JoarderKamruzzaman,"Survey of intrusion detection systems: techniques, datasets and challenges", Cyber Security Springer Open Access, 2019

[9] Iqbal H. Sarker, A. S. M. Kayes, ShahriarBadsha, HamedAlqahtani, Paul Watters, Alex Ng," Cybersecurity data science: an overview from machine learning perspective", Springer Open Access Journal on Big Data, 2020, PP. 1-29

[10] Ashfaq RAR, Wang X-Z, Huang JZ, Abbas H, He Y-L, " Fuzziness based semisupervised learning approach for instrusion detection system", Austrralian cyber security center threat report, 2017 Inf Sci 378:484-497

[11] https://www.assc.gov.au/publications /ACSC_Threat_Report_2017.pdf

[12] M. A. Alsheikh, S. Lin Niyato and H. P. Tan, " Machine learning in wiredless sensor networks: Algorithms, strategies and applications", IEEE communication Surveys and Tutorials, Vol. 16, no. 4, 2014, PP 1996-2018

[13] X. Xu, L. Zua and Z. Huang," Reinforcement learning algorithms with function approximation : Recent advances and applications", information Sciences, vol. 261, 2014, PP. 1-31

[14] W. Xing-zhu, " Network Intrusion Prediction Model based on RBF features classification", International Journal of Security and Its Applications, vol. 10, no. 4, 2016, PP. 241-248

[15] M. GhasemiGol and Ghaemi- Bafghi, " E-correlator: an entropy-based alert correlation system", Security and Communicaion Networks", vol. 8, no. 5, 2015, PP 822-836

[16] Ayei E. Ibor, Florence A. Oladeji and Olusoji B. Okunoye, " A Survey of cyber security Approaches for Attack detection, Prediction, and Prevention", International Journal of Security and Its applications, vol. 12, no. 4, 2018, PP. 15-28.

[17] P. S. Kenkre, A. Pai and Colaco, " Real time intrusion detection and prevention system", Proceedings of the 3rd International Conference on Frontiers of Intelligent Computing: Theory and Applications(FICTA), Springer, Cham 2014, PP 405-411

[18] Sunil Kumar, BhanuPratap Singh, Vinesh Kumar, " A Semantic Machine Learning Algorithm for Cyber Threat detection and Monitoring System", 3rd ICACCCN IEEE 2021, PP 1963-1967

[19] Kamran Shaukat, Suhuai Luo, Vijay Vardharajan, Ibrahim A. Hameed, and Min Xu, " A Survey on machine Learning Techniques for Cyber Security in the Last Decade", IEEE Access vol. 8, December 2020, PP 222310-222354